

ОРГАНИЗАЦИЯ ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ В ОБЛАСТИ МАТЕМАТИКИ НА ПРИМЕРЕ ИСПОЛЬЗОВАНИЯ ТЕЗАУРУСА ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Атаева О.М., ВЦ ФИЦ ИУ РАН

oli@ccas.ru

Серебряков В.А., д.ф.-м.н., профессор, ВЦ ФИЦ ИУ РАН

serebr@ccas.ru

Н.П.Тучкова, к.ф.-м.н., ВЦ ФИЦ ИУ РАН

tuchkova@ccas.ru

Аннотация. В фокусе предлагаемой работы предметные области, связанные с наукой, и их особенности. Сделана попытка выделения общих для них концепций. Одна из особенностей научных областей знаний заключается в том, что описываемые структуры данных подвержены частым изменениям. Обсуждается обобщенная модель научной предметной области, ее своеобразие, а также реализации в поисковых системах, отмечаются их отличия от классических подходов к поиску информации в научных массивах данных.

Ключевые слова: предметная область, научная предметная область, научная информация, научные знания, тезаурусы, адаптивная модель, организация научных знаний, цифровые библиотеки.

THE ORGANIZATION OF MATHEMATICAL KNOWLEDGE USING THE THESAURUS OF ORDINARY DIFFERENTIAL EQUATIONS AS AN EXAMPLE

Ataeva O.M., CC RAS

oli@ccas.ru

Serebryakov V.A., Grand PhD, professor, CC RAS

serebr@ccas.ru

N.P.Tuchkova, PhD, CC RAS

tuchkova@ccas.ru

Abstract. In this work, the subject areas related to science and their features are considered. An attempt has been made to single out common concepts for them. One of the features of the scientific fields of knowledge is that the described data structures are subject to frequent changes. The generalized model of the scientific subject domain and its originality are discussed as well as implementation in the search systems.

Keywords: subject area, scientific subject area, scientific information, scientific knowledge, thesauruses, adaptive model, organization of scientific knowledge, digital libraries.

Введение

С появлением парадигмы Semantic Web для формализации знаний в различных предметных областях стали активно использоваться онтологии. При этом данные предметной области имеют определенную структуру, зафиксированную в ней.

В предлагаемой работе рассматриваются предметные области, связанные с наукой и их особенности. Будет сделана попытка выделения общих концепций для их формальных описаний в базе знаний. Особенность этих областей заключается в том, что структура данных подвержена частым изменениям. Будем говорить об обобщенной модели научной предметной области и ее особенностях, реализациях в поисковых системах и отличий от классических подходов к поиску информации в научных массивах данных.

Актуальность такой работы связана с тем, что последние десятилетия объем информации лавинообразно увеличивается и это касается и научных областей. Продолжаются попытки построить формальные модели предметных областей. Резко возросло время, необходимое для поиска нужной информации и ее обзора.

Главной задачей создания описания обобщенного представления научных знаний для некоторой области является помощь экспертам в организации знаний и предоставления доступа к ней. При этом средство организации знаний должно быть достаточно универсальным и не требовать глубоких технических познаний.

Обобщенная модель научной предметной области

Построение обобщенной модели научной предметной области представляет подход, который делает упор на выделении таких метаданных, которые позволяют проектировать конкретные структуры данных для различных научных предметных областей и выявить общие подходы к управлению этими данными и их обработке. Это позволит построить общую модель *пространства научных знаний*, в рамках которой могут интегрироваться различные предметные области. С использованием этой модели возможна реализация семантических систем, способных гибко настраиваться под запросы конкретной предметной области. Одной из целей обобщенного подхода является упрощение доступа и восприятия больших и сложно структурированных объемов информации пользователем. Этот подход не является оптимальным для всех задач, решаемых в рамках некоторой научной предметной области, но, по крайней мере, позволяет структурировать имеющиеся знания на формальном уровне для дальнейшего использования.

Научная информация

Критерии научности информации строго не определены и на этот счет существуют различные точки зрения. Научная информация, по ГОСТ 7.0-99 [1]: логически организованная информация, получаемая в процессе научного познания и отображающая явления и законы природы, общества и мышления. Опираясь на это определение, можно выделить несколько основных свойств, которыми в совокупности обладает научная информация: истинность, интересобъективность и системность [2, 3].

Говоря о научной информации, имеет смысл разделять понятия научных *данных* и научных *ресурсов* в рамках научных предметных областей.

Научные ресурсы

В то же время, особенностью электронных и других источников научных данных является то, что, несмотря на сильное различие в назначении, они предоставляют похожие ресурсы для любой предметной области науки. То есть, информационные ресурсы в разных предметных областях представляются часто одними и теми же объектами: научные публикации, ученые, работающие в этой отрасли, научные организации, проекты, гранты, опыты, образцы, экспериментальные установки и другие. При этом непосредственно научные данные могут извлекаться из этих ресурсов.

Научные данные

Одновременно с построением обобщенной модели возникает необходимость ограничения ее в рамках конкретной предметной области науки. Для этого вводится набор понятий, используемых для описания этой предметной области. Соответствующие термины предметной области связывают с этими понятиями. Чаще всего эти термины организованы в виде некоторой таксономии с поддержкой связей между ними. Структура этой таксономии может варьироваться по сложности в зависимости от моделируемой области и представлять собой при необходимости полноценный тезаурус со всем богатством связей. В дальнейшем будем говорить о тезаурусах как о средстве организации понятий (знаний). Представленные в таком виде термины могут употребляться для обработки имеющихся ресурсов. При этом между понятиями и ресурсами возникают связи. Под научными данными

предметной области будем понимать совокупность понятий научной предметной области и выявленных связей между ними и ресурсами.

Научные знания

Совокупность научных ресурсов, их экземпляров, терминов тезауруса, всех явных и неявных связей между ними образуют общую картину научных знаний предметной области.

Самый простой набор правил выявления неявных связей определен в самом стандарте онтологического представления информации [4]. Онтологическое описание фактически является формальной основой для представления научных знаний и учета свойств интерсубъективности [5].

Модель данных

Формирование модели с перечисленными свойствами фактически подводит нас к построению онтологии, близкой, по сути, к высокоуровневым онтологиям для предметных областей науки. Такие онтологии описывают наиболее общие понятия, независимые от конкретной проблемы или области.

В этом смысле как нельзя лучше подходит адаптивная модель данных [6, 7] для описания научных ресурсов, которая позволяет не ограничиваться при разработке строго очерченным набором ресурсов. Такая модель данных подходит для определенного круга задач, решение которых реализуется на базе построения довольно сложных частных моделях. Применение адаптивной модели позволяет понизить сложность (размерность) как самой модели данных, так и разрабатываемых на их основе систем. Получаемые модели более абстрактны, состоят из меньшего количества понятий с более простыми связями и не привязаны к определенным предметным областям. Использование этой модели данных делает возможной динамическую трансформацию и интерпретацию модели данных в приложении, позволяя настраивать решения под определенную предметную область.

Поиск научной информации

Одним из несомненных преимуществ использования научной поисковой системы, построенной для некоторой предметной области науки, является ее качество поиска. В такого рода системах не встречается или встречается ничтожно мало случайных мусорных ресурсов и их источников. Это в свою очередь влияет на качество извлекаемой информации (достоверность, полнота, избыточность, надежность информационных источников), и сроки доставки ее до потребителя.

Очевидно, что при разработке общей идеологии нужно иметь в виду, что на глобальном уровне концептуализации должны присутствовать понятия, используя которые, можно описать структуру знаний любой научной предметной области. Выше мы разделили научные данные и научные ресурсы. В свою очередь для полноты необходимо ввести связи между данными и ресурсами. Перечислим основные понятия, необходимые для описания научных знаний: *Ресурс*, *Объект*, *Понятие*, *Связь понятиями*, *Тезаурус*, *Связь между понятиями и объектами*, *Правило*. Совокупность этих понятий задает обобщенную онтологию научной предметной области и вместе с данными составляют пространство научного знания рассматриваемой области.

С помощью этих понятий возможен переход на следующий уровень концептуализации – предметный. На предметном уровне для некоторой области происходит описание существенных для этой области понятий для научных ресурсов, данных и правил вывода новых знаний.

Пример реализации

Программной реализацией построенной модели описания научных предметных областей является система LibMeta [8]. Основная задача LibMeta – создание такой информационной системы для научных библиотек, которая могла бы учитывать все разнообразие различных типов ресурсов, которые могут в ней храниться и при этом поддерживать терминологическое описание практически любой предметной области. Одна из основных решаемых задач в контексте системы – это

обеспечение возможности интегрирования данных из различных источников. Фактически такая система представляет собой конструктор для создания цифровой научной библиотеки любой направленности и с произвольной моделью контента хранимых данных.

В качестве предметной области для демонстрации работы LibMeta с научной информацией рассматривалась область обыкновенных дифференциальных уравнений (ОДУ). Контент библиотеки определяется всего двумя типами научных ресурсов: Персона и Публикация. Для ограничения библиотеки в рамках этой предметной области науки был взят соответствующий тезаурус, построенный по стандарту ISO [9].

Основная задача состояла в том, чтобы после загрузки статей в LibMeta, основываясь на тезаурусе ОДУ, выделить те из них, которые относятся к данной предметной области. Для этого средствами системы были выделены связи между научными ресурсами, представленными публикациями, и терминами из тезауруса ОДУ.

Основное отличие предметной области математики, в частности ОДУ, от других научных предметных областей состоит в наличии формул. Причем формулы могут отличаться разнообразием записи, по сути представляя одну и ту же формулу. И если в настоящее время поиск по тексту стал массовым явлением, с поиском математических формул дело обстоит не столь гладко. Текстовый поиск, скорее всего, не даст нужного результата, поскольку формулы представляются в виде картинок или специализированных форматов [10].

Для поддержки формул было введено отдельное понятие *Формула*, которое позволяет хранить оригинальную строку формулы из источника, из которого она получена. Это понятие тесно связано отношениями с научными ресурсами и понятиями тезауруса рассматриваемой предметной области. Таким образом, получаем возможность построить сеть связей формулы с различными ресурсами и понятиями. Связи позволяют определять контекст формулы и области ее использования.

Выводы

В этой статье мы рассмотрели определение научной информации и выделили основные критерии и идеи для создания инструментария для ее описания и обработки. В качестве такого инструментария предлагается использовать информационную систему Libmeta, удовлетворяющую выдвинутым определениям.

Работая в качестве пилотного примера с областью обыкновенных дифференциальных уравнений, мы очертили круг понятий, необходимых для описания любой предметной области математики в общем виде, выделив в виде отдельного ресурса понятие формулы. Несомненно, эта область требует более детального описания, учитывая специфику научных ресурсов математики, но эта задача выходит за рамки статьи.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 17-07-00214-а.

Литература

1. ГОСТ 7.0-99 Межгосударственный стандарт ГОСТ 7.0-99 "Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения" (введен в действие постановлением Госстандарта РФ от 7 октября 1999 г. N 334-ст)
2. Губанов Николай Иванович, Губанов Николай Николаевич, Волков Андрей Эдуардович. "Критерии истинности и научности знания" Философия и общество, no. 3 (80), 2016, pp. 78-95.: <https://cyberleninka.ru/article/n/kriterii-istinnosti-i-nauchnosti-znaniya>
3. Ильин В. В., Калинин А. Т. Природа науки: Гносеологический анализ. М.: Высшая школа, 1985. – 230 с.
4. Semantic Web <http://www.w3.org/standards/semanticweb/>

5. Kumazawa Terukazu. "Toward knowledge structuring of sustainability science based on ontology engineering". Sustainability Science. 4: 99–116. doi:10.1007/s11625-008-0063-z. Retrieved 22 April 2015.
6. Leon Welick, Joseph W. Yoder, Rebecca Wirfs-Broc Adaptive Object-Model Builder – AdaptiveObjectModel.com, 2009 – <http://joeyoder.com/PDFs/04welicki.pdf>
7. Joseph W. Yoder, Federico Balaguer, Ralph Johnson Architecture and Design of Adaptive Object-Model – AdaptiveObjectModel.com, 2000, <http://www.adaptiveobjectmodel.com/OOPSLA2001/AOMIntriguingTechPaper.pdf>
8. Серебряков В.А., Атаева О.М. Информационная модель открытой персональной семантической библиотеки LibMeta // Труды XVIII Всероссийской научной конференции "Научный сервис в сети интернет". Новороссийск, 19 по 24 сентября 2016 г. ИПМ им. М.В. Келдыша РАН. С. 304-313.
9. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Тезаурус информационно-поисковый по предметной области «обыкновенные дифференциальные уравнения». М.: МАКС Пресс, 2005. 116 с.
10. Елизаров А. М., Липачев Е. К., Малахальцев М. А. Основы MathML. Представление математических текстов в Internet. Практическое руководство. Издательство Казанского математического общества, 2008.